AMENDMENTS TO THE SPECIFICATION

Please replace paragraph [0005], reference [6] of the original specification with the following rewritten paragraph:

--[0005] [6] M. A. Fischler and R. C. Bolles. Ransac random sample concensus: a paradigm for model ~~tting~~fitting with applications to image analysis and automated cartography. In Communications of the ACM, volume 26, 1981.—

Please replace paragraph [0005], reference [9] of the original specification with the following rewritten paragraph:

-- [0005] [9] E. Grimson, P. Viola, O.Faugeras, T. Lozano-Perez, T. Poggio, and S. Teller. A forest of ~~sensores~~sensors. In International Conference on Computer Vision, pages 45-51, 1997. --

Please replace paragraph [0005], reference [27] of the original specification with the following rewritten paragraph:

--[0005] [27] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, ~~Mosaic~~ Efficient Representations of Video Sequences and Their Applications. Signal Processing: Image Communication, special issue on Image and Video Semantics: Processing, Analysis, and Application, Vol. 8, No. 4, May 1996. –

Please replace paragraph [0027] of the original specification with the following rewritten paragraph:

--[0027] In the direct method, if there are no temporal changes in the scene, ~~i.e.,~~ then the temporal derivatives within the sequence are zero: $S_t \equiv 0$. Therefore, for any space-time point (x, y, t), the error term of Eq. (1) presented below reduces to:

$$\underbrace{err_{seq}\left(x,y,t;\vec{P}\right)}_{seq-to-seq} = S' - S + [u,v]\begin{bmatrix} S_x \\ S_y \end{bmatrix} = I' - I + [u,v]\begin{bmatrix} I_x \\ I_y \end{bmatrix} = \underbrace{err_{img}\left(x,y,t;\vec{P}\right)}_{img-to-img}$$

Where I(x,y)=S(x,y,t) is the image frame at time t. Therefore, the SSD function of Eq. (1) presented below reduces to the image-to-image alignment objective function, averaged over all frames.--

Please replace paragraph [0029] of the original specification with the following rewritten paragraph:

--[0029] The sequence-to-sequence paradigm is not limited only to direct methods, but can equally be implemented using feature-based methods. Feature-based methods first apply a local operator to detect singularity points on an image (e.g., Harris corner detector) [11]. Once two sets of singularity points are extracted, robust estimation methods such as RANSAC[6], LMS[7], etc. are used for finding corresponding points, and extracting the alignment parameters. –

Please replace paragraph [0030] of the original specification with the following rewritten paragraph:

--[0030] To address sequences instead of images, we extend the mining of a feature from feature point into feature trajectory. That is a trajectory of points representing its location on each frame within each sequence. Thus the second step will find correspondences between trajectories of points (the features x,y coordinates along the sequence). Note that in sequence-to-sequence alignment these trajectories do not necessarily belong to a moving object, but may include prominent features which ~~belongs~~ belong to a static object. This will produce a constant trajectory that is valid in any sense. –

Please replace paragraph [0031] of the original specification with the following rewritten paragraph:

--[0031] Feature based sequence-to-sequence alignment is a generalization of feature-based image-to-image alignment. If we consider a scene without moving objects, all trajectories will become trajectories of static objects, and the input is similar, thus the latter becomes identical to the first.—

Please replace paragraph [0036] of the original specification with the following rewritten paragraph:

--[0036] The paradigm of sequence-to-sequence alignment extends beyond any particular method. It can equally apply to feature-based matching across sequences, or other types of match measures (e.g., mutual information).—

Please replace paragraph [0048] of the original specification with the following rewritten paragraph:

--[0048] Fig. 7C is a pictorial illustrations of some possible alignments between the frames of Figs. 7A and 7B;--

Please replace paragraph [0069] of the original specification with the following rewritten paragraph:

--[0069] Fig. 6 is a diagram of a preferred method for subsampling and aligning image sequences according to a preferred embodiment of the present invention, where S'_0 is an original image sequence, S'_1 is subsampled ~~by~~from S'_0 as described herein, S'_2 is subsampled from S'_1 similarly, and so on.—

Please replace paragraph [0070] of the original specification with the following rewritten paragraph:

--[0070] Fig. 6 illustrates a preferred hierarchical spatio-temporal alignment framework. A volumetric pyramid is constructed for each input sequence, one for the reference sequence (on the right side), and one for the second sequence (on the left side). The spatio-temporal alignment estimator is applied iteratively at each level. It refines the approximation based on the residual misalignment between the reference volume and a warped version of the second volume (drawn as a skewed cube). The output of the current level is propagated to the next level to be used as an initial estimate.—

Please replace paragraph [0072] of the original specification with the following rewritten paragraph:

--[0072] Figs. 8A-8D illustrate spatio-temporal ambiguity in alignment when using only temporal information. A small airplane is crossing a scene viewed by two cameras. The airplane trajectory does not suffice to uniquely determine the alignment parameters. Arbitrary time shifts can be compensated by appropriate spatial translation along the airplane motion direction. Sequence-to-sequence alignment, on the other hand, can uniquely resolves this ambiguity, as it uses both the scene dynamics (the plane at different locations), and the scene appearance (the static ground). Note that spatial information alone does not suffice in this case either.—

Please replace paragraph [0074] of the original specification with the following rewritten paragraph:

--[0074] Fig. 11 illustrates a scene with moving objects. Lines 11(a) and 11(b) display 4 representative frames (100,200,300,400) from the reference and second sequences, respectively. The spatial misalignment is easily noticeable near image boundaries, where different static objects are visible in each sequence. The temporal misalignment is noticeable by comparing the position of the gate in frames 400. In the second sequence it is already open, while still closed in the reference sequence. Line ~~11~~11(c) displays superposition of the representative frames before spatio-temporal alignment. The superposition composes the red and blue bands from reference sequence with the green band from the second sequence. Line 11(d) displays superposition of corresponding frames after spatio-temporal alignment. The dark pink boundaries in (d) correspond to scene regions observed only by the reference camera. The dark green boundaries in (d) correspond to scene regions observed only by the second camera.—

Please replace paragraph [0076] of the original specification with the following rewritten paragraph:

--[0076] Fig. 13 illustrates a scene with non-rigid motion. Lines 13(a) and 13(b) display four representative frames (0,100,200,300) from the reference and second sequences, respectively. Line 13(c) displays superposition of the representative frames before spatio-temporal alignment. The spatial misalignment between the sequences is primarily due to scale differences in cameras focal length (i.e., differences in scale). The temporal misalignment is most evident in frames 300 of line 13(a) vs. 300 of line 13(b), where the wind blows the flag in reversed directions. Line 13(d) displays superposition of corresponding frames after spatio-temporal alignment.—

Please replace paragraph [0077] of the original specification with the following rewritten paragraph:

--[0077] Figs. 14-16 illustrate a scene which constantly changes its appearance. Figs. 14 and 15 display 10 frames (20,30, . . . ,110) from the reference and second sequences, respectively. It is difficult to tell the connection between the two sequences. The event in frames 90-110 in the reference sequence (Fig. 14), is the same as the event in frames 20-40 in the second sequence (Fig. 15). Fig. 16A displays superposition of the representative frames before spatio-temporal alignment. Fig. 16B displays superposition of corresponding frames after ~~ratio~~spatio-temporal alignment. Due to the scale difference there is an overlap between the two sequences only in the upper right region of every frame. Fireworks in the non-overlapping regions appear dark pink, as they were observed only by one camera. Fireworks in the overlapping regions appear white, as they should. The recovered temporal misalignment was approximately 66 frames.—

Please replace paragraph [0080] of the original specification with the following rewritten paragraph:

--[0080] For example, as shown in Fig. 1, a plurality of sequences of images are received, such as three sequences 50, 90 and 120 in the illustrated embodiment, captured by different image capturing devices typically imaging the same scene 180. In the illustrated embodiment, the image capturing devices are cameras I, II and III also designated by

reference numerals 20, 30 and 40 respectively. Each sequence, as shown, comprises an ordered multiplicity of images. For example, the sequence imaged by camera 20 is shown, for simplicity, to include three images 60, 70 and 80.—

Please replace paragraph [0081] of the original specification with the following rewritten paragraph:

--[0081] A particular advantage of a preferred embodiment of the invention as shown and described herein is that, as illustrated in Fig. 1, individual imaging processes, each of which have limitations, can, due to those limitations, represent an event so imperfectly as to be genuinely misleading. For example, as shown in Fig. 1, due to the insufficient temporal sampling employed by each of the three imaging devices 20, 30 and 40, none of the devices succeeds in correctly representing the S-shaped trajectory actually followed by the ball in the true scene ~~180~~160. The first camera 20, as shown in Figs. 2A-2B, perceives a straight-line trajectory because it images the ball only at positions 1, 6 and 11 which happen to fall roughly along a straight line. The second camera 30, as shown in Figs. 3A-3B, perceives a banana-shaped trajectory because it images the balls only at positions 2, 5 and 8 which happen to fall roughly along a banana-shaped curve. The third camera 40, as shown in Figs. 4A-4B, perceives an inverted banana-shaped trajectory because it images the balls only at positions 3, 7 and 11 which happen to fall roughly along an inverted banana-shaped curve.—

Please replace paragraph [0084] of the original specification with the following rewritten paragraph:

--[0084] In Fig. 5, the sequences are pictorially shown to be spatially aligned as evidenced by the three different orientations of the frames of type I, originating from camera I in Fig. 1, the frames of type II, originating from camera II, and the frames of type III, originating from camera III. As shown, the frames of type I are skewed such that their upper right hand corners ~~has~~ have been pivoted upward, the frames of type III are skewed such that their upper left hand corners ~~has~~ have been pivoted downward, and the

frames of type II are skewed intermediately between the frames of types I and III. The particular spatial misalignment illustrated pictorially in Fig. 1 is merely illustrative and is not intended to be limiting. Computational methods for effecting the alignment shown pictorially in Fig. 5 are described in detail herein.—

Please replace paragraph [0095] of the original specification with the following rewritten paragraph:

--[0095] A global constraint on $\vec{P}$ is obtained by minimizing the following SSD objective function:

$$ERR(\vec{P}) = \sum_{x,y,t} \left(S'(x,y,t) - S(x-u, y-v, t-w)\right)^2 \quad , \tag{1}$$

where $u = u(x,y,t:\vec{P}), v = v(x,y,t:\vec{P}), w = w(x,y,t:\vec{P})$ The parameter vector $\vec{P}$ is estimated e.g. using the Gauss-Newton minimization technique. To get a term which is explicit in the unknown parameters, linearize the term in Eq. 1 with respect to the parameter vector ~~P~~ $\underline{\vec{P}}$ to obtain:

$$ERP(\vec{P}) = \sum_{x,y,t} \left[S'(x,y,t) - S(x,y,t) + \nabla S(x,y,t) J_p \vec{P}\right]^2 \tag{2}$$

where $\nabla S = \left[S_x S_y S_t\right] = \left[\frac{\partial S}{\partial_x} \frac{\partial S}{\partial_y} \frac{\partial S}{\partial_t}\right]$ denotes a spatial-temporal gradient of the sequence S,

and J_p denotes the Jacobian matrix $J_p = \begin{bmatrix} \frac{\partial x'}{\partial P_1} & & \frac{\partial x'}{\partial P_n} \\ \frac{\partial y'}{\partial P_1} & \cdots & \frac{\partial y'}{\partial P_n} \\ \frac{\partial t'}{\partial P_1} & & \frac{\partial t'}{\partial P_n} \end{bmatrix}$

P is estimated by least-squares minimization. We solve the following "normal equations":

$$\sum_{x,y,t} J_P \nabla S(x,y,t)(J_P \nabla S(x,y,t))^T \vec{P} = \sum_{x,y,t} [S'(x,y,t) - S(x,y,t)] J_P \nabla S(x,y,t) \quad (3)$$

or in short notations:

$$\sum_{x,y,t} J_P \nabla S (J_P \nabla S)^T \vec{P} = \sum_{x,y,t} [S' - S] J_P \nabla S \quad (4)$$

A different linearization (with respect to (x,y,t)) is possible as well:

$$e(x,y,t;\vec{P}) = S'(x,y,t) - S(x,y,t) + [uvw] \nabla S(x,y,t) \quad (\cancel{3}\underline{5})$$

and $\nabla S = [S_x S_y S_t] = \left[\frac{\partial S}{\partial_x} \frac{\partial S}{\partial_y} \frac{\partial S}{\partial_t}\right]$ denotes a spatio-temporal gradient of the sequence S. Eq. ($\cancel{3}\underline{5}$) directly relates the unknown displacements (u, v ,w) to measurable brightness variations within the sequence. To allow for large spatio-temporal displacements (u, v, w), the minimization of Equations (1) ~~or~~ (3) <u>or (5)</u> is done within an iterative-warp coarse-to-fine framework as described herein.

Please replace paragraph [0098] of the original specification with the following rewritten paragraph:

--[0098] **Model 1**: *2D spatial affine transformation and 1D temporal affine transformation.* The spatial 2D affine model is obtained by setting the third row of H to be: H3 = [0, 0, 1]. Therefore, for 2D spatial affine and 1D temporal affine transformations, the unknown parameters are:

$\underline{\vec{P} = [h_{11}\ h_{12}\ h_{13}\ h_{21}\ h_{22}\ h_{23}\ d_1\ d_2]}$ i.e., eight unknowns. The individual voxel error of Eq. ($\cancel{3}\underline{5}$) becomes:

$$e(x,y,t;\vec{P}) = S' - S + [(H_1\vec{p} - x)(H_1\vec{p} - y)(d_1 t + d_2)] \nabla S,$$

which in linear in all unknown parameters.

Model 2: *2D spatial projective transformation and a temporal offset.* In this case, w(t)=d (d is a real number, i.e., could be a sub-frame shift), and $\vec{P} = [h_{11}, h_{12}\ h_{13}\ h_{21}\ h_{22}\ h_{23}\ h_{31}\ h_{32}\ h_{33}\ d]$

Each spatio-temporal "voxel" (x, y, t) provides one constraint:

$$e(x, y, t; \vec{P}) = S' - S + \left[\left(\frac{H_1\vec{p}}{H_3\vec{p}} - x\right)\left(\frac{H_2\vec{p}}{H_3\vec{p}} - y\right)(d_1 t + d_2)\right]\nabla S' \qquad \cancel{(4)}\ (6)$$

The 2D projective transformation is not linear in the unknown parameters, and hence preferably undergoes some additional manipulation. To overcome this non-linearity, Eq. ~~(4)~~ (6) is multiplied by the denominator $(H_3\vec{P})$, and renormalized with its current estimate from the last iteration, leading to a slightly different error term: -

$$e_{new}(x, y, t; \vec{P}) = \frac{H_3\vec{P}}{\hat{H}_3\vec{P}} \cdot e_{old}(x, y, t; \vec{P}), \qquad (\cancel{5}7)\text{—}$$

where $\hat{H}_3$ is the current estimate of H_3 in the iterative process, and e_{old} is as defined in Eq. ~~(4)~~ (6).—

Please replace paragraph [0099] of the original specification with the following rewritten paragraph:

--[0099] Let $\hat{H}$ and $\hat{d}$ be the current estimates of H and d, respectively. Substituting $H = \hat{H} + \delta H$ and $d = \hat{d} + \delta d$ into Eq. (~~5~~7), and neglecting high-order terms, leads to a new error term, which is linear in all unknown parameters (δH and δd). In addition to second order terms (e.g., $\delta H\ \delta d$), the first order term $\hat{d}\delta H_3$ is also negligible and can be ignored.--

Please replace paragraph [0105] of the original specification with the following rewritten paragraph:

-- [0105] In our experiments, two different interlaced CCD cameras (mounted on tripods) were used for sequence acquisition. No synchronization what so ever was used. Typical sequence length is several hundreds of frames. Lines (a)-(d) in Fig. 11 shows a scene with a car driving in a parking lot. The two input sequences line 11(a) and line 11(b) were taken from two different windows of a tall building. Line 11(c) displays superposition of representative frames, generated by mixing the red and blue bands from the reference

sequence with the green band from the second sequence. This demonstrates the initial misalignment between the two sequences, both in time and in space. Note the different timing of the gate being lifted (temporal misalignment), and misalignment of static scene parts, such as the parked car or the bushes (spatial misalignment). Line 11(d) shows the superposition after applying spatio-temporal alignment. The second sequence was spatio-temporally warped towards the reference sequence according to the computed parameters. The recovered spatial affine transformation indicated a translation on the order of a 1/5 of the image size, a small rotation, a small scaling, and a small skew (due to different aspect ratios of the two cameras). The recovered temporal shift was 46.63 frames. Therefore, opposite fields at a distance of 46 frames were mixed together when applying the color superposition.--

Please replace paragraph [0106] of the original specification with the following rewritten paragraph:

--[0106] In Fig. 12, the sequences (a)-(d) illustrate that dynamic information cues are not restricted to independent object motion. A light source was brightened and then dimmed down, resulting in observable illumination variations in the scene. The cameras were imaging a picture on a wall from significantly different viewing angles, inducing a significant perspective distortion. Line (a) and line (b) show a few representative frames from two sequences of several hundred frames each. The effects of illumination are particularly evident in the upper left corner of the image. Note the difference in illumination in frame 200 of the two sequences--frame 200 in line 12(a) and frame 200 in line 12(b). Line 12(c) shows a superposition of the representative frames from both sequences before spatio-temporal alignment. Line 12(d) shows superposition of corresponding frames after spatio-temporal alignment. The correctness of the temporal alignment is evident from the hue in the upper left corner of frame 200, which is pink before alignment (frame 200 in line 12(c)) and white after temporal alignment (frame 200 in line 12(d)).The accuracy of the recovered temporal offset (21.32 frames) was verified (up to 0.1 frame time) against the ground truth. The verification was implemented by imaging a small object (a tennis ball) that moves very fast. The objects ~~was~~were viewed

by three fields only (not included in the part that was used ~~to~~for alignment). The tennis ball location enables us to verify manually that correct field-to field temporal corresponds. Furthermore, the phase differences of these locations (3 in each sequences) produce sub field accuracy "ground truth". We manually distinguish between 5 phases ~~of~~ $\left\{0,\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5}\right\}$ of field time. Therefore we can verify our results up to 0.1 frame time. This suggests that the temporal offset was a component of 0.6 of field time. --

Please replace paragraph [0107] of the original specification with the following rewritten paragraph:

--[0107] In Fig. 13 the sequences (a)-(d) illustrate a case where the dynamic changes within the sequence are due to non-rigid motion (a flag blowing in the wind). Line 13(a) and line 13(b) show two representative frames out of several hundred. Line 13(c) shows a superposition of the representative frames from both sequences before spatio-temporal alignment. Line 13(d) shows superposition of corresponding frames after spatio-temporal alignment. The recovered temporal ~~o set~~offset was 31.43 frames. Image-to-image alignment performs poorly in this case, even when applied to temporally corresponding frames, as there is not enough spatial information in many of the individual frames. This is shown in Fig. 10. We applied image-to-image alignment to all temporally corresponding pairs of fields, (odd fields from one camera with even fields from the second camera as the computed time shift (31.4) is closer to 31.5 than the 31.0) Only 55% of corresponding frames converged to accurate spatial alignment. The other 45% suffered from noticeable spatial misalignment. A few representative frames (out of the 45% of misaligned pairs) are shown in Fig. 10, line (a). These pairs were well aligned by sequence-to-sequence alignment (Fig. 10, line (b)).--

Please replace paragraph [0108] of the original specification with the following rewritten paragraph:

--[0108] Figs. 14-16 illustrates that temporal changes may include changes in appearance

of the entire scene. The sequences show an explosion of fireworks. The fireworks change their appearance (size, shape, color and brightness) drastically throughout the sequence. Figs. 14 and 15 show ten representative frames from two sequences of a few hundreds frames each. Frames 20-110 are displayed from the both sequences. The event in frames 90-110 in the reference sequence (Fig. 14), is the same as the event shown in frames 20-40 in the second sequence (Fig. 15). Line 16(a) displays superposition of four representative frames (80-110) before applying spatio-temporal alignment. The fireworks appear green and pink, due to the superposition of the different bands from different sequences (red and blue from one sequence and green from the other). The artificial colors are due to the mixture of misaligned fireworks with dark background from the other sequence. Line 16(b) displays superposition of the same five representative frames after applying spatio-temporal alignment. The fireworks are now white in the overlapping image regions, as they should be, implying good spatio-temporal alignment.—

Please replace paragraph [0109] of the original specification with the following rewritten paragraph:

--[0109] The above results were mainly qualitative. To quantify the expected accuracy of the method we applied several experiments ~~were~~ where the exact ground truth alignment was known. First we warped a sequence using ~~a~~ known spatio-temporal parameters, applied our method to the warped and original sequence and compared the extracted parameters with the known ones. This produced highly accurate results. Less than 0.01 frame time temporal error, and less than 0.02 pixels spatial error. The accurate results are due to the fact that the source and warped sequences are highly correlated. The only difference in corresponding "voxels" gray level is as a results of the tri-linear interpolation used when ~~of the~~ warping the second sequence. To create a test ~~were~~ where the noise is less correlated we split a sequence into its two fields. The two field" sequences are related by known temporal and spatial parameters a temporal shift of 0.5 frame time, and temporal shift of 0.5 pixel along the Y axis. Note~~,~~ that in this case the data comes from the same camera, but from completely different sets of pixels (odd rows

in one sequence and even rows in the other sequence). We repeated the experiment several (10) times using different sequences and different spatial models (affine, projective). In all cases the temporal error was smaller than~~then~~ 0.02 frame time. (i.e., the recovered time shift was between 0.48 and 0.52). The recovery~~ed~~ in the Y-translation was smaller then 0.03 pixel (i.e., the recovered Y-shift was between 0.47 and 0.53 pixel)~~.~~ and the overall Euclidean error ~~all~~ over the image was bounded by 0.1 pixels. To include error that results from using two cameras, we applied this test to pairs of sequences ~~form~~from different cameras.—

Please replace paragraph [0110] of the original specification with the following rewritten paragraph:

--[0110] Each sequence was split into two sequences of odd and even fields. In this case the ground truth is not given but the relative change is known. That is if the time shift between odd sequences from the first camera reference cameras is δt then the time shift between odd sequences from the first camera and even sequences from the reference camera should be δt +0.5, and the same holds for spatial alignment. This also was performed several times and in all cases the temporal error was bounded by 0.05 frame time and the spatial error was bounded by 0.1 pixel.—